UNIVERSITI TEKNOLOGI MARA

PROFILING NETWORK TRAFFIC OF SULTAN IDRIS SHAH BUILDING (BSIS) USING DATA MINING TECHNIQUE

RUSMAWATI BINTI ISHAK

Dissertation submitted in partial fulfillment
of the requirements for the degree of
Master of Science in Computing Networking

Faculty of Computer and Mathematical Science

July 2018

ABSTRACT

With the exponential of information technology and data growth, a rapid increase in the request to the system and workload may cause severe congestion. A high performance of BSIS network infrastructure and services that is capable to serve thousands or millions of traffic for BSIS users faster, more reliable, lower response time and will bring good experience to the online users. Hence, in order to make it reliable to provide good online services, we need to identify what kind of application, characteristic trend and behaviour of BSIS users in network infrastructure. Thus, in this research we focus on analysing BSIS network traffic to find a pattern that cause of slow connection during peak hours and non-peak hours that will come out with network profiling information. The network profiling will help the agency to give better solution in managing network and security policy and infrastructure as well as to give better experienced in online activities request by users. This research implement unsupervised data mining approach to analyse the network traffic trend of BSIS. Orange is a tool that being used in implementing K-Means Clustering Algorithm that could be seen as the most suitable solution to find a network trend pattern of user accessing the Internet and to produce with profiling network. The result showed that the K-Means method can perform clustering with three (3) cluster in which describe based on high, medium and low number of hits towards the protocol services and unique IP address. Based on the result, most higher traffic pass through the network that consume of bandwidth usage are DNS (UDP) with total of 74,819 hits, MySQL with total of 40,691 and HTTP service with 59,784 hits record. Based on the trend shows more activities done during the holiday and the attacking activity also frequently happen during the weekend or public holiday. Data mining process lead to reveal the information gather for profiling purposes and identify type of traffic passing through the BSIS network. Furthermore, the outcome of this study can be a recommendation of managing or shaping the bandwidth usage and strengthen the security policy as well as help the organization to manage the network and assist the connectivity issue faced by internal and online user.

ACKNOWLEDGEMENT

Alhamdulillah praises and thanks to Allah S.W.T because of His Almighty and His utmost blessing, I was able to finish this dissertation within the time duration given. It is a long and challenging journey to face in order to complete this dissertation.

Firstly, I would like to express my gratitude and special thanks to my supervisor Dr Siti Arpah Ahmad for her guidance and advice from time to time until completion of this research. I would like to give my sincere appreciation to all the lecturers who taught me a lot during my studies here also to my second supervisor Mr. Adzhar Abdul Kadir. Besides that, I also would like to extend my gratitude to my examiners for assisting me in preparing a dissertation report.

Special thanks to my industrial superior, Mrs Nurul Elliza Jasmin for giving me the opportunity and permission to do my research study also to my colleagues and my dearest friends for their contribution and helping me to complete this research.

Finally, special appreciation to my beloved parents [redacted] and Mr Ishak Ismail for their understanding, advice and support towards my studies. Thank you very much, may Allah bless all of us.

TABLE OF CONTENT

Page

CHAPTER ONE
INTRODUCTION

This chapter provides the background and rationale for the study. It provides significant details of the study, the issues and problem related to this research.

1.1 Research Background

Nowadays, the Internet is a necessity in obtaining and delivering information easily and quickly without using paper especially to organizations offering online services to their customers. The discussion today is no longer focus on infrastructure provided only, but the key thing that is required nowadays is about the performance on providing services. Basically, TCP/IP is a well-known protocol use in Information Technology. In World Wide Web (WWW) using standard rules in communication known as HTTP or HTTPS by which browsers fetch documents and other resources from web servers. All transactions will pass through the network infrastructure. In network and security management, the use of high speed Internet are increasing, in line with rapid development of application and network, as well as mobile devices. Table 1.1 shows the survey done by MCMC to determine the percentage of Internet users in Malaysia for the years 2015 and 2016. Based on the statistics show there are 0.4 million of Internet user increased in year 2016 compared to 2015 (Internet Users Survey, MCMC, 2017).

Table 1.1
Percentage of Internet Users in Malaysia

Year	Percentage of Internet Users	Number of Internet Users
2016	76.9%	24.5 million users
2015	77.6%	24.1 million users

(Source: Internet Users Survey MCMC, 2017)

Rapid development of Internet technology transformed our way of getting information, entertainment, socializing, and participating in online activities. Figure